

51

Int. Cl. 2:

H 04 B 3/04
G 10 L 1/00

19 BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENTAMT



IV 11/00 C7 1/1
DT 25 06 771 A1

11

Offenlegungsschrift 25 06 771

21

Aktenzeichen: P 25 06 771.4

22

Anmeldetag: 18. 2. 75

43

Offenlegungstag: 19. 8. 76

30

Unionspriorität:

32 33 31 -

54

Bezeichnung: Verfahren zur Verbesserung der Sprechererkennung

71

Anmelder: Philips Patentverwaltung GmbH, 2000 Hamburg

72

Erfinder: Bunge, Ernst, Dipl.-Ing., 2000 Hamburg; Rothgordt, Ulf, Dipl.-Phys. Dr., 2000 Norderstedt

BEST AVAILABLE COPY

IV 11/00 C7 1/1

2506771

PHILIPS PATENTVERWALTUNG GMBH, 2 Hamburg 1, Steindamm 94

Verfahren zur Verbesserung der Sprechererkennung

Die Erfindung betrifft ein Verfahren zur Verbesserung der Sprechererkennung bei Übertragung des zu erkennenden Sprachsignals über eine Übertragungsstrecke mit linearen Verzerrungen zu dem Erkennungsgerät.

Um mit einem automatischen rechnergesteuerten Sprechererkennungssystem eine hohe Erkennungsrate zu erzielen, ist es notwendig, die zu erkennenden Sprachproben bzw. Sprachsignale dem Rechner immer in gleichbleibender Qualität zur Verfügung zu stellen. Bei Sprachübertragung beispielsweise per Telefon ist jedoch bei jeder Telefonverbindung die unterschiedliche Übertragungsfunktion der Strecke Mikrofonskapsel - Telefonleitung - Telefonadapter zwischen dem eigentlichen Sprachsignal und dem Signal, das dem Rechner zugeführt wird, unvermeidlich zwischengeschaltet. Der Einfluß dieser jeweils wechselnden Übertragungsfunktion kann so groß sein, daß eine zuverlässige Sprechererkennung sehr erschwert bzw. unmöglich gemacht wird.

PHD 75/025 Po/Sa

609834/0578

- 2 -

ORIGINAL INSPECTED

Aufgabe der Erfindung ist es, ein Verfahren zur Verbesserung der Sprechererkennung anzugeben, bei dem Auswirkungen durch die linearen Verzerrungen des Sprachsignals auf der Übertragungsstrecke weitgehend vermieden werden. Diese Aufgabe löst die Erfindung durch die im Kennzeichen des Hauptanspruchs angegebenen Merkmale. Das vorbestimmte Signal kann ein festgelegter Text, also ein Testsatz sein, der vor Beginn der eigentlichen Übertragung von dem zu erkennenden Sprecher besprochen wird. Eine andere Möglichkeit besteht darin, dem Mikrofon auf der Sprecherseite ein Sinussignal in Form eines von einem Lautsprecher erzeugten Tones mit konstanter Amplitude und einer sich stetig von der einen bis zur anderen Grenzfrequenz der Übertragungsstrecke ändernden Frequenz zuzuführen. Dafür ist zwar ein gewisser, wenn auch geringer apparativer Aufwand notwendig, jedoch werden mit dieser Methode die Eigenschaften der Übertragungsstrecke objektiver erfaßt.

Die Korrektur des übertragenen Sprachsignals kann dadurch erfolgen, daß mit den Werten der im Erkennungsgerät bestimmten Übertragungsfunktion ein elektronisch einstellbares Filter entsprechend eingestellt wird, oder daß bei einem Erkennungsgerät, das mittels einer Filterbank das Sprachsignal in die einzelnen Spektralanteile zerlegt, die Ausgänge der Filterbank entsprechend gewichtet werden. Die Korrektur durch die Übertragungsfunktion kann jedoch auch implizit erfolgen, indem das Fourier-Spektrum des übertragenen Sprachsignals mit dem Quotienten aus dem Fourier-Spektrum

des gespeicherten vorbestimmten Signals und des übertragenen vorbestimmten Signals multipliziert wird.

Das Sprachsignal des zu erkennenden Sprechers sei durch seine Amplituden-Zeit-Funktion charakterisiert. Wird dieses Sprachsignal über eine Telefonleitung übertragen, so erhält man als Ausgangssignal beim Erkennungsgerät eine Funktion, die durch die Übertragungseigenschaften des Telefonübertragungsweges verändert worden ist. Um zu zeigen, wie aus dem Ausgangssignal das originale Sprachsignal näherungsweise wiedergewonnen werden kann, werden beide Signale fouriertransformiert, so daß sich folgende Beziehung zwischen beiden Signalen ergibt:

$$F_{1T}(\omega) = F_1(\omega) \cdot A_T(\omega)$$

Die unbekannte Übertragungsfunktion $A_T(\omega)$ ist dann

$$A_T(\omega) = \frac{F_{1T}(\omega)}{F_1(\omega)}$$

Darin ist die Fouriertransformierte $F_1(\omega)$ des originalen Sprachsignals ebenfalls noch unbekannt.

In dem Erkennungssystem ist nun ein gesprochener Testsatz bzw. dessen Eigenschaften gespeichert. Wenn der gleiche Testsatz vom gleichen Sprecher noch einmal gesprochen wird, so ist er, innerhalb der normalen Reproduzierbarkeitsgrenzen, am Ort des Sprechers dem originalen Testsatz sehr ähnlich. Wenn also der noch einmal gesprochene Testsatz im Erkennungssystem mit dem originalen Testsatz verglichen wird, kann daraus also die unbekannte Übertragungs-

609834/0578

funktion $A_T(\omega)$ näherungsweise bestimmt werden:

$$A_T(\omega) \approx \frac{F_{2, 2T}(\omega)}{F_{2, 1}(\omega)}$$

Darin ist $F_{2, 2T}(\omega)$ die Fouriertransformierte des noch einmal gesprochenen und übertragenen Testsatzes und $F_{2, 1}(\omega)$ die Fouriertransformierte des gespeicherten Testsatzes ist.

Mit der somit zumindest näherungsweise bestimmten Übertragungsfunktion kann nun in bekannter Weise die Inversfilterung des zu erkennenden Sprachsignals im Rechner vorgenommen werden, um den Einfluß der Telefonstrecke auf die Sicherheit der Sprechererkennung zu reduzieren. Wenn insbesondere das Spektrum des Sprachsignals untersucht wird, kann die Übertragungsfunktion $A_T(\omega)$ gleichsam implizit berücksichtigt werden, um aus dem Spektrum $F_{1T}(\omega)$ das originale Sprachsignal $F_1(\omega)$ zu rekonstruieren:

$$F_1(\omega) = \frac{F_{1T}(\omega)}{A_T(\omega)}$$

$$F_1(\omega) = \frac{F_{1T}(\omega)}{F_{2, 2T}(\omega)} \cdot F_{21}(\omega)$$

Eine andere Möglichkeit besteht darin, daß nicht ein Testsatz, sondern ein Sinussignal konstanter Amplitude mit variabler Frequenz über die Telefonleitung geschickt wird. Die dafür notwendige

Anordnung kann beispielsweise aus einem Sägezahngenerator, einem Sinusgenerator, einem Verstärkerbaustein und einem kleinen Lautsprecher bestehen. Der Sägezahngenerator steuert die Frequenz des Sinusgenerators vorzugsweise linear über den Übertragungsbereich von etwa 100 Hz bis 5 Khz. Diese Sinusspannung wird über den Verstärker einem Miniaturlautsprecher zugeführt, der direkt vor die Mikrofonkapsel des Telefons gehalten wird. Bei Verwendung von integrierten Schaltkreisen und einem Miniaturlautsprecher läßt sich diese Anordnung in der Größe einer Telefonkapsel aufbauen, die vor jeder Übertragung eines zu erkennenden Sprachsignals vor die Telefonmikrofonkapsel gehalten wird. Auf der Empfängerseite braucht dann nur noch von dem empfangenen Signal eine Spektralanalyse gemacht zu werden, um die Übertragungsfunktion zu ermitteln und damit die Sprechererkennung nahezu unabhängig von dem Frequenzgang der Telefonleitung durchzuführen.

2506771

Patentansprüche:

=====

1. Verfahren zur Verbesserung der Sprechererkennung bei Übertragung des zu erkennenden Sprachsignals über eine Übertragungsstrecke mit linearen Verzerrungen zu dem Erkennungsgerät, dadurch gekennzeichnet, daß vor der Übertragung des Sprachsignals über die Übertragungsstrecke ein vorbestimmtes Signal übertragen wird, daß in dem Erkennungsgerät das vorbestimmte Signal bzw. dessen charakteristischen Merkmale gespeichert sind und mit dem bzw. denen des übertragenen vorbestimmten Signals verglichen werden, und daß aus dem Vergleich eine Übertragungsfunktion bestimmt wird, mit der das folgende übertragene, zu erkennende Sprachsignal vor der Erkennung korrigiert wird.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das vorbestimmte Signal ein festgelegter Text ist.
3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das vorbestimmte Signal ein Sinussignal mit konstanter Amplitude und einer sich stetig von der einen bis zur anderen Grenzfrequenz der Übertragungsstrecke ändernde Frequenz ist.
4. Verfahren nach Anspruch 2 oder 3, dadurch gekennzeichnet, daß das Fourier-Spektrum des übertragenen Sprachsignals mit dem Quotienten aus den Fourier-Spektren des gespeicherten vorbestimmten Signals und des übertragenen vorbestimmten Signals multipliziert wird.

BEST AVAILABLE COPY

609834/0578

THIS PAGE BLANK (USPTO)